

CLAIMS

1. A method for evaluating the text content of a document database with respect to a document population, comprising the steps of:

(a) providing a computer system having a user interface with a display;

5 (b) gathering documents from said database into said system;

(c) normalizing said gathered documents;

(d) fingerprinting said gathered documents;

(e) determining a text criteria with respect to said document population;

10 (f) forming a net comprising at least two nodes associated by at least one interaction and displayable at said display as two or more spaced apart nodes connected by an interaction;

(g) loading said text criteria into at least one of said nodes;

15 (h) for each document of said database, calculating its geometric relative distance from a said node to derive one or more node attractors;

(i) displaying said net at said display in combination with one or more document symbols each representing a said document located in correspondence with said calculated relative distance;

20 (j) visually examining the display of said net and document symbols; and

(k) determining from said document symbol locations at said display those documents, if any, which are more likely to correspond with said text criteria.

25 2. The method of claim 1 in which:

said step (f) forming a net provides for the display of said net as having said nodes defined as circles and said interaction defined as a line extending between said circles.

30 3. The method of claim 1 in which:

said step (g) loads said text criteria into a positive designated one of said nodes.

4. The method of claim 1 in which:
said step (f) forms said net as comprising a said positive designated node and a null designated node connected by a said interaction.
5. The method of claim 1 in which:
said step (e) determines said
text criteria as criteria document textual material; and
said step (g) comprises the steps:
(g1) normalizing said criteria document textual material; and
(g2) fingerprinting the normalized criteria document textual material.
6. The method of claim 1 in which:
said step (e) determines a positive text criteria and a negative text criteria with respect to said document population;
said step (f) forms a net comprising one or more positive designated nodes, one or more negative designated nodes and one or more interactions;
said step (g) loads said positive text criteria into said one or more positive designated nodes, and said negative text criteria into said one or more negative designated nodes; and
said step (h) calculates, for each document of said database its geometric relative distance from both said positive designated node and said negative designated node.
7. The method of claim 1 in which:
said step (i) displays said one or more document symbols as squares.
8. The method of claim 1 including the steps:
(l) retrieving the identification of those documents resulting from the determination of step (k);
(m) viewing one or more of the documents identified in step (l) and determining the quality of the match thereof with said step (e) text criteria.
9. The method of claim 8 further comprising the steps:

(n) identifying a new text criteria as a result of a said step (m) determination of an insufficient said quality of said match;

(o) adding the identified new text criteria to the step (g) text criteria loaded into said positive designated one of said nodes; and

5 (p) reiterating said steps (h) through (m).

10. The method of claim 8 further comprising the steps:

(q) subsequent to said step (m), identifying and viewing at said display a list of features common to those documents the identification of which was
10 retrieved in step (l);

(r) identifying a new text criteria in correspondence with said step (q) identification and viewing at said display of said features common to those documents the identification of which was retrieved in step (l);

(s) adding the identified new text criteria to the step (q) text criteria
15 loaded into said positive designated one of said nodes; and

(t) reiterating said steps (h) through step (m).

11. The method of claim 1 in which:

said step (k) further comprises the steps:

(k1) determining additional text criteria where said document symbol
20 locations are not likely to correspond with said text criteria determined at step (e); and

(k2) adding said additional text criteria to said text criteria determined at said step (e).

25 12. The method of claim 8 in which:

said step (l) is carried out by drawing at said display of said net a boundary defining a region of said symbols.

13. A method for evaluating the text content of a document database with
30 respect to a population of documents. comprising the steps of:

(a) providing a computer system having a user interface with a display;

(b) forming one or more nets each comprising at least two nodes associated by at least one interaction, one or more said nodes representing an evaluation criteria, said one or more nets being viewable at said display;

5 (c) treating said documents to have an attribute value and calculating for each document a geometric relative distance with respect to a said node criteria and displaying corresponding document symbols at said display;

(d) delimiting at said display a first region of said document symbols;

10 (e) delimiting at said display a second region of said document symbols;

(f) selecting a said document attribute to be correlated and the criteria for establishing an attribute value match;

(g) determining the presence of one or more document attribute value match pairs as correlations between said first and second regions; and

15 (h) displaying said correlations at said display.

14. The method of claim 13 in which:

said step (d) provides a said first region extending over more than one said net; and includes the step:

20 (d1) mapping said first region to a first document set by selecting the union or intersection of documents on different nets.

15. The method of claim 13 in which:

25 said step (e) provides a said second region extending over more than one said net; and including the step:

(e1) mapping said second region to a second document set by selecting the union or intersection of documents on different nets.

16. The method of claim 13 in which:

30 said step (f) selects said criteria for establishing an attribute value match by defining an attribute value tolerance.

17. The method of claim 16 in which:

said step (g) determines the presence of a document attribute match pair by determining whether the attribute value of a document in said first region is equal to the attribute value of a document in said second region within said attribute value tolerance.

5

18. The method of claim 13 in which:

said step (d) is carried out by providing a computer generated line or lines visible at said display.

10

19. The method of claim 13 in which:

said step (e) is carried out by providing a computer generated line or lines visible at said display.

20. The method of claim 13 in which:

15

said step (h) is carried out by providing visible line at said display connecting two said symbols and representing said correlation.

21. The method of claim 13 in which:

said step (f) selects said document attribute or document identification;

20

and

said step (g) identifies the same document in each said first and second region as a said correlation.

22. A method for searching the text content of a document database with respect to a population of documents, comprising the steps of:

25

(a) providing a computer system having a user interface with a display;

(b) identifying the population of documents to be searched;

(c) normalizing the documents of the identified population with the

30

steps comprising;

(c1) selecting character sequences that will separate words,

(c2) determining to either retain or eliminate punctuation characters,

(c3) setting regular expressions that will characterize numbers,

(c4) setting case behavior,

- (c5) setting an offset and factor for numeric class,
- (c6) converting a document of said identified population to a character sequence,
- (c7) accessing the words, or punctuation characters, W of said character sequences,
- (c8) for each accessed W which is a number, converting such number into a sequence of word numbers, WN, and normalizing said word numbers for fingerprinting,
- (c9) marking the position and length of each W or normalized word number WN,
- (c10) for each W or normalized WN, completing said normalization by reiterating steps (c8) and (c9);
- (d) fingerprinting said normalized documents;
- (e) forming one or more nets, each comprising at least two nodes, one or more said nodes representing an evaluation criteria, said one or more nets exhibiting two or more spaced apart nodes connected by one or more interactions;
- (f) for each normalized document, calculating its geometric relative distance from a said node;
- (g) displaying said one or more nets at said display in combination with one or more document symbols representing a said document located in correspondence with said calculated relative distance; and
- determining from said document symbol locations at said display, if any, those documents which are more likely to correspond with said evaluation criteria.

25

23. The method of claim 22 in which said step (c8) further comprises the steps:

- (c81) convert any date characterized word number WN to a float or integer,
- (c8.2) applying an offset and factor to the word number WN to derive X,
- (c8.3) set the range, R,
- (c8.4) calculate the quantity $T = (\log_{10}X)/R$,

(c8.5) successively decrementing the value of range, R and calculating the quantity, T until R is equal to zero,

(c8.6) designating S as the position of a significant numeral in X,

5 (c8.7) assigning each successive quantity T to a corresponding successive position S to derive the first component of normalized word number WN; and

(c8.8) subsequent to said step (c8.7), assigning each successive said numeral in X to a corresponding successive position S to derive a second component of said normalized word number, WN.

10

24. The method of claim 23 in which:

said step (c8.3) further comprises the step: (c8.3.1) setting the precision P of the normalized word number WN,

15 said step (c8.8) is carried out until the number of said successive positions S deriving said second component equals the value of said precision, R.